

0590  
1124

Page 1 of 7

#6



OIKE

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/837,602

DATE: 02/10/2002

TIME: 13:38:33

Input Set : N:\Crf3\RULE60\09837602.txt

Output Set: N:\CRF3\02102002\I837602.raw

4 <110> APPLICANT: Petrini, John H.J.  
5 Morgan, William Franklin  
6 Maser, Richard Scott  
7 Carney, James Patrick  
10 <120> TITLE OF INVENTION: DNA Encoding A DNA Repair Protein  
13 <130> FILE REFERENCE: 800.019US1  
15 <140> CURRENT APPLICATION NUMBER: US 09/837,602  
17 <141> CURRENT FILING DATE: 2001-04-18  
19 <150> PRIOR APPLICATION NUMBER: US 09/067,641  
21 <151> PRIOR FILING DATE: 1998-04-27  
23 <160> NUMBER OF SEQ ID NOS: 24  
25 <170> SOFTWARE: FastSEQ for Windows Version 3.0  
27 <210> SEQ ID NO: 1  
28 <211> LENGTH: 4403  
29 <212> TYPE: DNA  
30 <213> ORGANISM: Homo sapiens  
32 <400> SEQUENCE: 1

ENTERED

33	ttcggcacga	ggcgcggttg	cacgtcggcc	ccagccctga	ggagccggac	cgatgtggaa	60
34	actgctgccc	gccgcgggcc	cggcaggagg	agaaccatac	agacttttga	ctggcggttga	120
35	gtacgttggt	ggaaggaaaa	actgtgccat	tctaattgaa	aatgatcagt	cgatcagccg	180
36	aaatcatgct	gtgttaactg	ctaacttttc	tgtaaccaac	ctgagtcaaa	cagatgaaat	240
37	cctgtatttg	acattaaaaa	ataattctaa	gtatgggtacc	tttggttaatg	aggaaaaaat	300
38	gcagaatggc	ttttcccgaa	ctttgaagtc	gggggatggg	attacttttg	gagtgtttgg	360
39	aagtaaatc	agaatagagt	atgagccttt	ggttgcattg	tcttcttggt	tagatgtctc	420
40	tgggaaaact	gctttaaatc	aagctatatt	gcaacttgga	ggatttactg	taaacaattg	480
41	gacagaagaa	tgacttcacc	ttgtcatggg	atcagtgaag	gttaccatta	aaacaatatg	540
42	tgactcatt	tgtggacgtc	caattgtaaa	gccagaatat	tttactgaat	tcttgaaagc	600
43	agttcagtc	aagaagcagc	ctccacaaat	tgaaagtgtt	taccacctc	ttgatgaacc	660
44	atctattgga	agtaaaaatg	ttgatctgtc	aggacggcag	gaaagaaaac	aaatcttcaa	720
45	agggaaaaca	tttataattt	tgaatgcca	acagcataag	aaattgagtt	ccgcagttgt	780
46	ctttggaggt	gggggaagcta	ggttgataac	agaagagaat	gaagaagaac	ataatttctt	840
47	tttggctccg	ggaacgtgtg	ttgttgatac	aggaataaca	aactcacaga	ccttaattcc	900
48	tgactgtcag	aagaaatgga	ttcagtcagt	aatggatatg	ctccaaaggc	aaggctcttag	960
49	acctattcct	gaagcagaaa	ttggattggc	ggtgattttc	atgactacaa	agaattactg	1020
50	tgatcctcag	ggccatccca	gtacaggatt	aaagacaaca	actccaggac	caagcctttc	1080
51	acaaggcgtg	tcagttgatg	aaaaactaat	gccaaagcgc	ccagtgaaca	ctacaacata	1140
52	cgtagctgac	acagaatcag	agcaagcaga	tacatgggat	ttgagtgaag	ggccaaaaga	1200
53	aatcaaaagt	tccaaaatg	aacaaaaatt	cagaatgctt	tcacaagacg	caccactgt	1260
54	aaaggagtcc	tgcaaaaaca	gctctaataa	taatatgatg	gtatcaaaata	ctttggctaa	1320
55	gatgagaatc	ccaaaactatc	agcttttcacc	aactaaattg	ccaagtataa	ataaaaagtaa	1380
56	agatagggct	tctcagcagc	agcagaccaa	ctccatcaga	aactactttc	agccgtctac	1440
57	caaaaaaagg	gaaagggatg	aagaaaatca	agaaatgtct	tcatgcaaat	cagcaagaat	1500
58	agaaacgtct	tggttctctt	tagaacaac	acaacctgct	acacctcat	tgtggaaaaa	1560

## RAW SEQUENCE LISTING

DATE: 02/10/2002

PATENT APPLICATION: US/09/837,602

TIME: 13:38:33

Input Set : N:\Crf3\RULE60\09837602.txt

Output Set: N:\CRF3\02102002\I837602.raw

```

59 taaggagcag catctatctg agaatgagcc tgtggacaca aactcagaca ataacttatt 1620
60 tacagataca gatttaaaat ctattgtgaa aaattctgcc agtaaatctc atgctgcaga 1680
61 aaagctaaga tcaaaataaaa aaagggaaat ggatgatgtg gccatagaag atgaagtatt 1740
62 ggaacagtta ttcaaggaca caaaaccaga gttagaaatt gatgtgaaag ttcaaaaaca 1800
63 ggaggaagat gtcaatgtta gaaaaaggcc aaggatggat atagaaacaa atgacacttt 1860
64 cagtgatgaa gcagtaccag aaagtagcaa aatatctcaa gaaaatgaaa ttgggaagaa 1920
65 acgtgaactc aaggaagact cactatggtc agctaaagaa atatctaaca atgacaaact 1980
66 tcaggatgat agtgagatgc ttccaaaaaa gctgttattg actgaattta gatcactggt 2040
67 gattaaaaac tctacttcca gaaatccgct tggcataaat gatgattatg gtcaactaaa 2100
68 aaatttcaag aaattcaaaa aggtcacata tcctggagca ggaaaacttc cacacatcat 2160
69 tggaggatca gatctaatag ctcatcatgc tcgaaagaat acagaactag aagagtggct 2220
70 aaggcaggaa atggagggtac aaaatcaaca tgcaaaagaa gagtctcttg ctgatgatct 2280
71 ttttagatac aatccttatt taaaaaggag aagataactg aggattttta aaagaagcca 2340
72 tggaaaaact tcctagtaag catctacttc aggccaacaa ggttatatga atatatagt 2400
73 tatagaagcg atttaagtta caatgtttta tggcctaaat ttattaaata aaatgcacaa 2460
74 aactttgatt cttttgtatg taacaattgt ttgtyctggt ttcaggcttt gtcattgcat 2520
75 ctttttttca tttttaaatg tgttttgttt attaaatagt taatatagtc acagttcaaa 2580
76 attctaaatr tacgtaagggt aaaggactaa agtcaccctt ccaccattgt cctagctact 2640
77 tattttttaa taatttccta cacaaatgat agcataacat atgcagtgtt ctacaccttg 2700
78 cttttttact tagtaagatt aaaaattata ggaatatcaa tataatgttt ttaatatttt 2760
79 ttcttttcca ttatgctgta gtcttaccta aactctggtg atccaaacaa aatggcttca 2820
80 gtggtgcaga tgtcacctac atgttattct agtactagaa actgaagacc atgtggagac 2880
81 ttcacaaac atgggtttag ttttcaccag aatggaaaga cctgtacccc ttttgggtgg 2940
82 tcttactgag ctgggtgggt gtctgttttg agcttattta ggtcctagt tttcctactt 3000
83 ataaagtaga aatggtgaga ttgttttctt tttctacckt aaaggagat ggtaagaaac 3060
84 aatgaatgtc ttttttcaaa ctttattgac aagtgtttt caagtctgtg ttcaaaaata 3120
85 tattcatgta cctgtgatcc agcaagaagg gagtccagt caagagtcac tacaactgat 3180
86 tagttgttta gagaatgaga aatggaacag tgaggaaatgg aggccatatt tccatgactt 3240
87 cccttgtaaa cagaagcaac agaagggaca agaggctggc ctctacatca ctctcacctt 3300
88 ccaaactctg tggaaagtga tctacttgcc agaaccaaat taacttactt ccaagttctg 3360
89 gctgcttgca ggtggaactc cagctgcaag ggagttaggg aaatgaagggt ctttttttaa 3420
90 aagcttctca gccttcctag ggaacagaaa ttgggtgagc caatctgcaa tttctactac 3480
91 aggcattgag accagttaga ttattgaaat attatagaga gttatgaaca cttaaattat 3540
92 gatagtggta tgacattgga tagaacatgg gatactttag aagtagaatt gacagggcat 3600
93 attagttagt gaaatggagt catttgagtc tyttaatagc catgtatcat aattaccaag 3660
94 tgaagctggt ggaacatatg gtctccattt tacagttaag gaatataatg gacagattaa 3720
95 tattgttytc tgtcatgccc acaatccctt tctaaggaaag actgccctac tatagcagtt 3780
96 tttatatttg tcaatttatg aatataatga atgaggagtt ctggtacctc ctgtctttac 3840
97 aaatattggg tgtgtccag tatttttccc tttttaaccm ttcccaattc ggggtgtgag 3900
98 gtggatgttt ccatttgggt ttttaattgt atatccctga tagctataat tgggtcatag 3960
99 aaattcttta tacattctag atgcaagtct cttgycggat atacgtattg agatattaca 4020
100 cctagtctgt ggcttgactg ttttctttat gtcttttgat gaatagaagt tttaaatttt 4080
101 gacaaggcca aatttatatt tttcttttgt ttgatatttt ttctctccaa ttttaaccca 4140
102 agatttcaga tattctgctc tattatataa actttatatt ttatatttg tgatctacct 4200
103 tgaattgata tgtatgttgt gaattatgga tcagggttct ttttttcccc catacaagta 4260
104 tcaagtcaat gtaacactgt ttattgaaag aattatcctt tcctcattaa attaccttgc 4320
105 caattagtaa aaaaatcaatt aacctatmar mmmrrrggat ccactagttc tagagcggcc 4380
106 gccaccgcgg tggagctcca gct 4403
108 <210> SEQ ID NO: 2

```

## RAW SEQUENCE LISTING

DATE: 02/10/2002

PATENT APPLICATION: US/09/837,602

TIME: 13:38:33

Input Set : N:\Crf3\RULE60\09837602.txt

Output Set: N:\CRF3\02102002\I837602.raw

```

109 <211> LENGTH: 754
110 <212> TYPE: PRT
111 <213> ORGANISM: Homo sapiens
113 <400> SEQUENCE: 2
114 Met Trp Lys Leu Leu Pro Ala Ala Gly Pro Ala Gly Gly Glu Pro Tyr
115 1 5 10 15
116 Arg Leu Leu Thr Gly Val Glu Tyr Val Val Gly Arg Lys Asn Cys Ala
117 20 25 30
118 Ile Leu Ile Glu Asn Asp Gln Ser Ile Ser Arg Asn His Ala Val Leu
119 35 40 45
120 Thr Ala Asn Phe Ser Val Thr Asn Leu Ser Gln Thr Asp Glu Ile Pro
121 50 55 60
122 Val Leu Thr Leu Lys Asp Asn Ser Lys Tyr Gly Thr Phe Val Asn Glu
123 65 70 75 80
124 Glu Lys Met Gln Asn Gly Phe Ser Arg Thr Leu Lys Ser Gly Asp Gly
125 85 90 95
126 Ile Thr Phe Gly Val Phe Gly Ser Lys Phe Arg Ile Glu Tyr Glu Pro
127 100 105 110
128 Leu Val Ala Cys Ser Ser Cys Leu Asp Val Ser Gly Lys Thr Ala Leu
129 115 120 125
130 Asn Gln Ala Ile Leu Gln Leu Gly Gly Phe Thr Val Asn Asn Trp Thr
131 130 135 140
132 Glu Glu Cys Thr His Leu Val Met Val Ser Val Lys Val Thr Ile Lys
133 145 150 155 160
134 Thr Ile Cys Ala Leu Ile Cys Gly Arg Pro Ile Val Lys Pro Glu Tyr
135 165 170 175
136 Phe Thr Glu Phe Leu Lys Ala Val Gln Ser Lys Lys Gln Pro Pro Gln
137 180 185 190
138 Ile Glu Ser Phe Tyr Pro Pro Leu Asp Glu Pro Ser Ile Gly Ser Lys
139 195 200 205
140 Asn Val Asp Leu Ser Gly Arg Gln Glu Arg Lys Gln Ile Phe Lys Gly
141 210 215 220
142 Lys Thr Phe Ile Phe Leu Asn Ala Lys Gln His Lys Lys Leu Ser Ser
143 225 230 235 240
144 Ala Val Val Phe Gly Gly Gly Glu Ala Arg Leu Ile Thr Glu Glu Asn
145 245 250 255
146 Glu Glu Glu His Asn Phe Phe Leu Ala Pro Gly Thr Cys Val Val Asp
147 260 265 270
148 Thr Gly Ile Thr Asn Ser Gln Thr Leu Ile Pro Asp Cys Gln Lys Lys
149 275 280 285
150 Trp Ile Gln Ser Ile Met Asp Met Leu Gln Arg Gln Gly Leu Arg Pro
151 290 295 300
152 Ile Pro Glu Ala Glu Ile Gly Leu Ala Val Ile Phe Met Thr Thr Lys
153 305 310 315 320
154 Asn Tyr Cys Asp Pro Gln Gly His Pro Ser Thr Gly Leu Lys Thr Thr
155 325 330 335
156 Thr Pro Gly Pro Ser Leu Ser Gln Gly Val Ser Val Asp Glu Lys Leu
157 340 345 350
158 Met Pro Ser Ala Pro Val Asn Thr Thr Thr Tyr Val Ala Asp Thr Glu

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/837,602

DATE: 02/10/2002

TIME: 13:38:33

Input Set : N:\Crf3\RULE60\09837602.txt

Output Set: N:\CRF3\02102002\I837602.raw

```

159          355          360          365
160 Ser Glu Gln Ala Asp Thr Trp Asp Leu Ser Glu Arg Pro Lys Glu Ile
161          370          375          380
162 Lys Val Ser Lys Met Glu Gln Lys Phe Arg Met Leu Ser Gln Asp Ala
163 385          390          395          400
164 Pro Thr Val Lys Glu Ser Cys Lys Thr Ser Ser Asn Asn Asn Ser Met
165          405          410          415
166 Val Ser Asn Thr Leu Ala Lys Met Arg Ile Pro Asn Tyr Gln Leu Ser
167          420          425          430
168 Pro Thr Lys Leu Pro Ser Ile Asn Lys Ser Lys Asp Arg Ala Ser Gln
169          435          440          445
170 Gln Gln Gln Thr Asn Ser Ile Arg Asn Tyr Phe Gln Pro Ser Thr Lys
171          450          455          460
172 Lys Arg Glu Arg Asp Glu Glu Asn Gln Glu Met Ser Ser Cys Lys Ser
173 465          470          475          480
174 Ala Arg Ile Glu Thr Ser Cys Ser Leu Leu Glu Gln Thr Gln Pro Ala
175          485          490          495
176 Thr Pro Ser Leu Trp Lys Asn Lys Glu Gln His Leu Ser Glu Asn Glu
177          500          505          510
178 Pro Val Asp Thr Asn Ser Asp Asn Asn Leu Phe Thr Asp Thr Asp Leu
179          515          520          525
180 Lys Ser Ile Val Lys Asn Ser Ala Ser Lys Ser His Ala Ala Glu Lys
181          530          535          540
182 Leu Arg Ser Asn Lys Lys Arg Glu Met Asp Asp Val Ala Ile Glu Asp
183 545          550          555          560
184 Glu Val Leu Glu Gln Leu Phe Lys Asp Thr Lys Pro Glu Leu Glu Ile
185          565          570          575
186 Asp Val Lys Val Gln Lys Gln Glu Glu Asp Val Asn Val Arg Lys Arg
187          580          585          590
188 Pro Arg Met Asp Ile Glu Thr Asn Asp Thr Phe Ser Asp Glu Ala Val
189          595          600          605
190 Pro Glu Ser Ser Lys Ile Ser Gln Glu Asn Glu Ile Gly Lys Lys Arg
191          610          615          620
192 Glu Leu Lys Glu Asp Ser Leu Trp Ser Ala Lys Glu Ile Ser Asn Asn
193 625          630          635          640
194 Asp Lys Leu Gln Asp Asp Ser Glu Met Leu Pro Lys Lys Leu Leu Leu
195          645          650          655
196 Thr Glu Phe Arg Ser Leu Val Ile Lys Asn Ser Thr Ser Arg Asn Pro
197          660          665          670
198 Ser Gly Ile Asn Asp Asp Tyr Gly Gln Leu Lys Asn Phe Lys Lys Phe
199          675          680          685
200 Lys Lys Val Thr Tyr Pro Gly Ala Gly Lys Leu Pro His Ile Ile Gly
201          690          695          700
202 Gly Ser Asp Leu Ile Ala His His Ala Arg Lys Asn Thr Glu Leu Glu
203 705          710          715          720
204 Glu Trp Leu Arg Gln Glu Met Glu Val Gln Asn Gln His Ala Lys Glu
205          725          730          735
206 Glu Ser Leu Ala Asp Asp Leu Phe Arg Tyr Asn Pro Tyr Leu Lys Arg
207          740          745          750

```

## RAW SEQUENCE LISTING

DATE: 02/10/2002

PATENT APPLICATION: US/09/837,602

TIME: 13:38:33

Input Set : N:\Crf3\RULE60\09837602.txt

Output Set: N:\CRF3\02102002\I837602.raw

```

208 Arg Arg
211 <210> SEQ ID NO: 3
212 <211> LENGTH: 87
213 <212> TYPE: PRT
214 <213> ORGANISM: Homo sapiens
216 <220> FEATURE:
217 <221> NAME/KEY: UNSURE
218 <222> LOCATION: (48)...(48)
219 <223> OTHER INFORMATION: Unsure
221 <400> SEQUENCE: 3
222 Tyr Val Val Gly Arg Lys Asn Cys Ala Ile Leu Ile Glu Asn Asp Gln
223 1 5 10 15
224 Ser Ile Ser Arg Asn His Ala Val Leu Thr Ala Asn Phe Ser Val Thr
225 20 25 30
W--> 226 Asn Leu Ser Gln Thr Asp Glu Ile Pro Val Leu Thr Leu Lys Asn Xaa
227 35 40 45
228 Lys Tyr Gly Thr Phe Val Asn Glu Glu Lys Met Gln Asn Gly Phe Ser
229 50 55 60
230 Arg Thr Leu Lys Ser Val Asp Gly Ile Thr Phe Gly Val Phe Gly Ser
231 65 70 75 80
232 Lys Phe Arg Ile Glu Tyr Glu
233 85
235 <210> SEQ ID NO: 4
236 <211> LENGTH: 87
237 <212> TYPE: PRT
238 <213> ORGANISM: Homo sapiens
240 <400> SEQUENCE: 4
241 Tyr Ser Ile Gly Arg Ser Ser Lys Asn Pro Leu Ile Ile Lys Asn Asp
242 1 5 10 15
243 Lys Ser Ile Ser Arg Gln His Ile Thr Phe Lys Trp Glu Ile Asn Asn
244 20 25 30
245 Ser Ser Asp Leu Lys His Ser Ser Leu Cys Leu Val Asn Lys Gly Lys
246 35 40 45
247 Leu Thr Ser Leu Asn Lys Lys Phe Met Lys Val Gly Glu Thr Phe Thr
248 50 55 60
249 Ile Asn Ala Ser Cys Val Leu Lys Ser Thr Ile Glu Leu Gly Thr Thr
250 65 70 75 80
251 Pro Ile Arg Ile Glu Phe Glu
252 85
254 <210> SEQ ID NO: 5
255 <211> LENGTH: 13
256 <212> TYPE: PRT
257 <213> ORGANISM: Homo sapiens
259 <400> SEQUENCE: 5
260 Asn Pro Ser Gly Leu Asn Asp Asp Tyr Gly Gln Leu Lys
261 1 5 10
263 <210> SEQ ID NO: 6
264 <211> LENGTH: 680
265 <212> TYPE: PRT

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/837,602

DATE: 02/10/2002

TIME: 13:38:34

Input Set : N:\Crf3\RULE60\09837602.txt

Output Set: N:\CRF3\02102002\I837602.raw

L:226 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3